

[Click to Take the FREE Crash-Course](#)

How to Choose a Feature Selection Method For Machine Learning

by Jason Brownlee on November 27, 2019 in [Data Preparation](#)

[Tweet](#)[Share](#)[Share](#)

Last Updated on May 14, 2020

Feature selection is the process of reducing the number of input variables when developing a predictive model.

It is desirable to reduce the number of input variables to both reduce the computational cost of modeling and, in some cases, to improve the performance of the model.

Statistical-based feature selection methods involve evaluating the relationship between each input variable and the target variable using statistics and selecting those input variables that have the strongest relationship with the target variable. These methods can be fast and effective, although the choice of statistical measures depends on the data type of both the input and output variables.

As such, it can be challenging for a machine learning practitioner to select an appropriate statistical measure for a dataset when performing filter-based feature selection.

In this post, you will discover how to choose statistical measures for filter-based feature selection with numerical and categorical data.

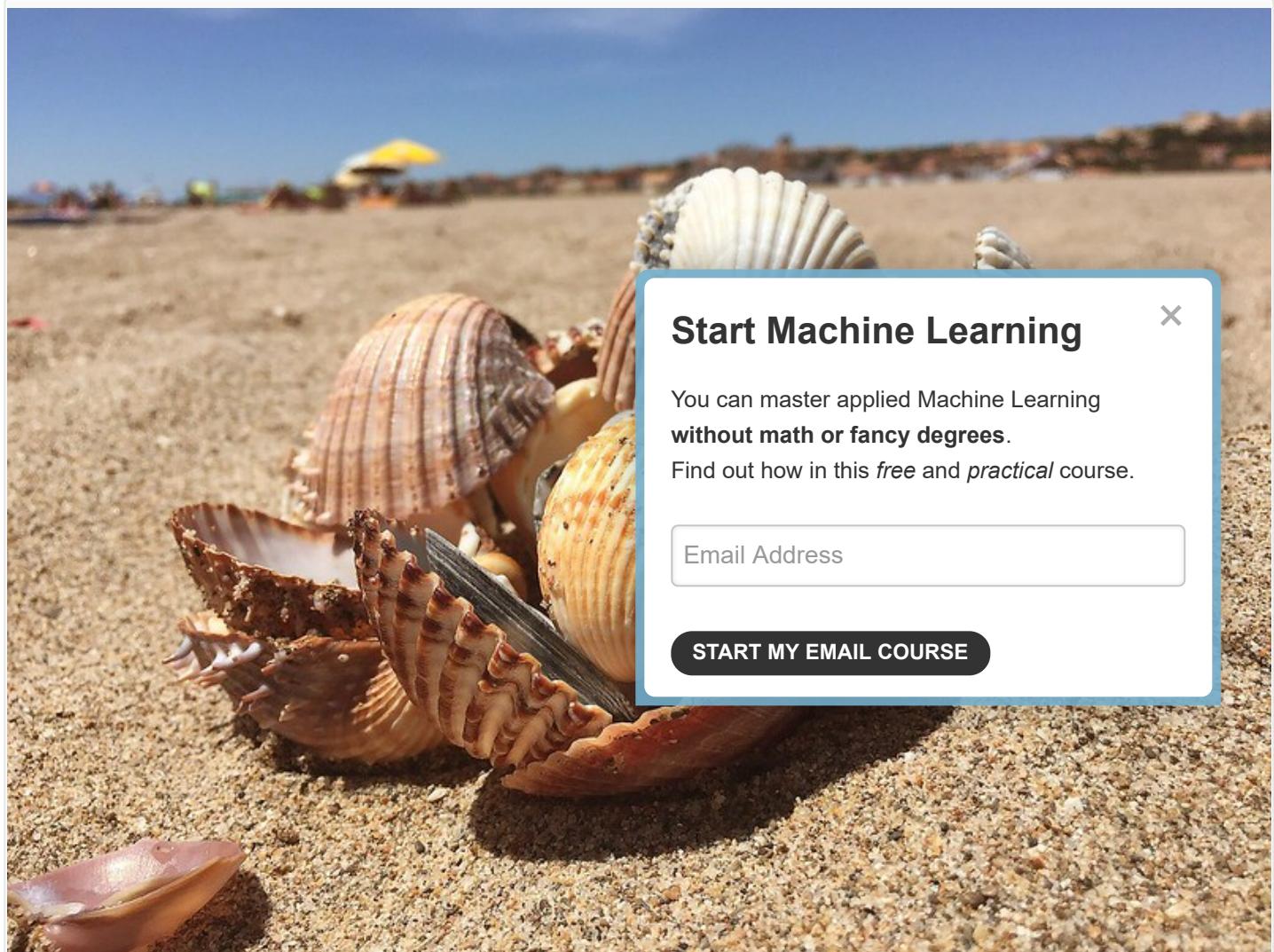
After reading this post, you will know:

- There are two main types of feature selection techniques: supervised and unsupervised, and supervised methods may be divided into wrapper, filter and intrinsic.
- Filter-based feature selection methods use statistical measures to score the correlation or dependence between input variables that can be filtered to choose the most relevant features.
- Statistical measures for feature selection must be carefully chosen based on the data type of the input variable and the output or response variable.

[Start Machine Learning](#)

Let's get started.

- **Update Nov/2019:** Added some worked examples for classification and regression.
- **Update May/2020:** Expanded and added references. Added pictures.



Start Machine Learning X

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

How to Develop a Probabilistic Model of Breast Cancer Patient Survival

Photo by Tanja-Milfoil, some rights reserved.

Overview

This tutorial is divided into 4 parts; they are:

1. Feature Selection Methods
2. Statistics for Filter Feature Selection Methods
 1. Numerical Input, Numerical Output
 2. Numerical Input, Categorical Output
 3. Categorical Input, Numerical Output
 4. Categorical Input, Categorical Output
3. Tips and Tricks for Feature Selection
 1. Correlation Statistics

Start Machine Learning ^

2. Selection Method
3. Transform Variables
4. What Is the Best Method?
4. Worked Examples
 1. Regression Feature Selection
 2. Classification Feature Selection

1. Feature Selection Methods

Feature selection methods are intended to reduce the number of input variables to those that are believed to be most useful to a model in order to predict the target variable.

“ Feature selection is primarily focused on removing variables that are not useful to the model.

— Page 488, [Applied Predictive Modeling](#), 2013.

Some predictive modeling problems have a large number of input variables. This can lead to slow training of models and require a large amount of system resources. In addition, the performance of some models can degrade when including input variables that are not useful to the model.

“ Many models, especially those based on regression, estimate a separate set of regression parameters for every term in the model. Because of this, the presence of non-informative variables can add uncertainty to the predictions and reduce the overall effectiveness of the model.

— Page 488, [Applied Predictive Modeling](#), 2013.

One way to think about feature selection methods are in terms of **supervised** and **unsupervised** methods.

“ An important distinction to be made in feature selection is that of supervised and unsupervised methods. When the outcome is ignored during the elimination of predictors, the technique is unsupervised.

— Page 488, [Applied Predictive Modeling](#), 2013.

The difference has to do with whether features are selected based on the target variable or not. Unsupervised feature selection techniques ignores the target variable, such as methods that remove redundant variables using correlation. Supervised feature selection techniques use the target variable, such as methods that remove irrelevant variables..

Another way to consider the mechanism used to select features which may be divided into **wrapper** and **filter** methods. These methods are almost always supervised and are evaluated based on the performance

[Start Machine Learning](#)

of a resulting model on a hold out dataset.

Wrapper feature selection methods create many models with different subsets of input features and select those features that result in the best performing model according to a performance metric. These methods are unconcerned with the variable types, although they can be computationally expensive. [RFE](#) is a good example of a wrapper feature selection method.

“ *Wrapper methods evaluate multiple models using procedures that add and/or remove predictors to find the optimal combination that maximizes model performance.*

— Page 490, [Applied Predictive Modeling](#), 2013.

Filter feature selection methods use statistical techniques to calculate the relevance of each predictor variable and the target variable, and these scores are used to select a subset of the most relevant variables that will be used in the model.

“ *Filter methods evaluate the relevance of the predictors and subsequently model only the predictors that pass the filter.*

— Page 490, [Applied Predictive Modeling](#), 2013.

Finally, there are some machine learning algorithms that automatically select a subset of the input variables for learning the model. We might refer to these techniques as **intrinsic** feature selection methods.

Start Machine Learning

You can master applied Machine Learning without **math or fancy degrees**.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

“ *... some models contain built-in feature selection, meaning that the model will only include predictors that help maximize accuracy. In these cases, the model can pick and choose which representation of the data is best.*

— Page 28, [Applied Predictive Modeling](#), 2013.

This includes algorithms such as penalized regression models like Lasso and decision trees, including ensembles of decision trees like random forest.

“ *Some models are naturally resistant to non-informative predictors. Tree- and rule-based models, MARS and the lasso, for example, intrinsically conduct feature selection.*

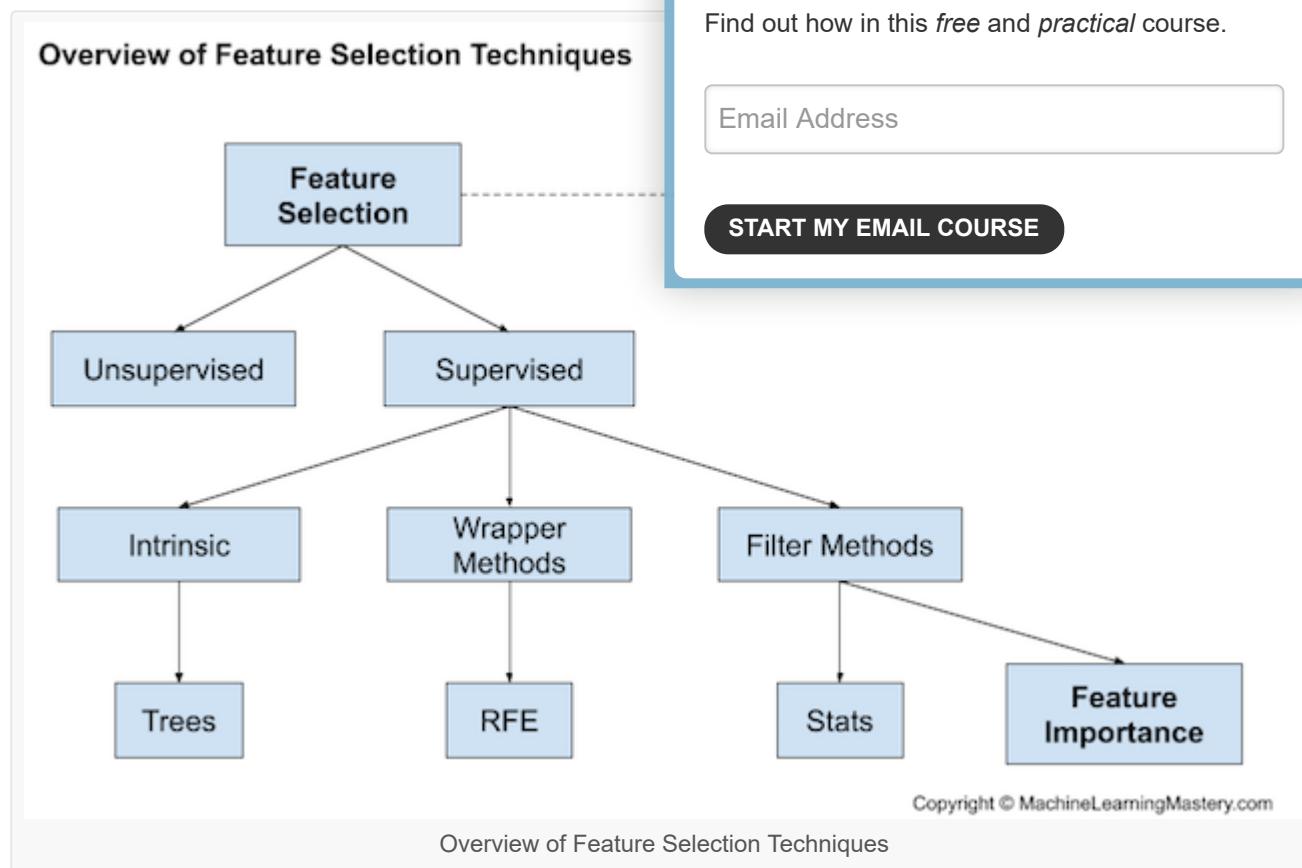
— Page 487, [Applied Predictive Modeling](#), 2013.

Feature selection is also related to [dimensionally reduction](#) techniques in that both methods seek fewer input variables to a predictive model. The difference is that feature selection select features to keep or remove from the dataset, whereas dimensionality reduction create a projection of the data resulting in entirely new input features. As such, dimensionality reduction is an alternate to feature selection rather than a type of feature selection.

We can summarize feature selection as follows.

- **Feature Selection:** Select a subset of input features from the dataset.
 - **Unsupervised:** Do not use the target variable (e.g. remove redundant variables).
 - Correlation
 - **Supervised:** Use the target variable (e.g. remove irrelevant variables).
 - **Wrapper:** Search for well-performing subsets of features.
 - RFE
 - **Filter:** Select subsets of features based on their relationship with the target.
 - Statistical Methods
 - Feature Importance Methods
 - **Intrinsic:** Algorithms that perform automatic feature selection during training.
 - Decision Trees
- **Dimensionality Reduction:** Project input data into a lower-dimensional space.

The image below provides a summary of this hierarchy.



In the next section, we will review some of the statistical measures that may be used for filter-based feature selection with different input and output variable data types.

2. Statistics for Filter-Based Feature Selection Methods

It is common to use correlation type statistical measures between input and output variables as the basis for filter feature selection.

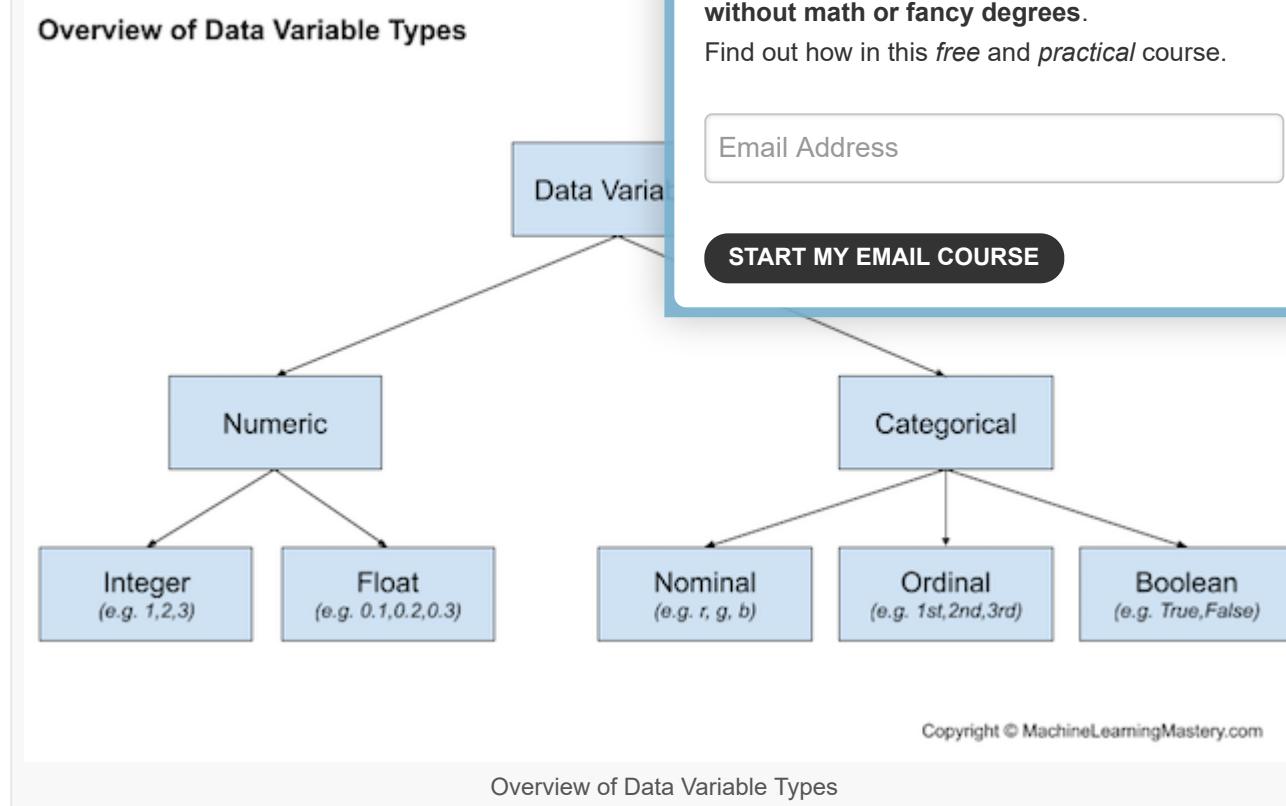
[Start Machine Learning](#)

As such, the choice of statistical measures is highly dependent upon the variable data types.

Common data types include numerical (such as height) and categorical (such as a label), although each may be further subdivided such as integer and floating point for numerical variables, and boolean, ordinal, or nominal for categorical variables.

Common input variable data types:

- **Numerical Variables**
 - Integer Variables.
 - Floating Point Variables.
- **Categorical Variables.**
 - Boolean Variables (dichotomous).
 - Ordinal Variables.
 - Nominal Variables.



The more that is known about the data type of a variable, the easier it is to choose an appropriate statistical measure for a filter-based feature selection method.

In this section, we will consider two broad categories of variable types: numerical and categorical; also, the two main groups of variables to consider: input and output.

Input variables are those that are provided as input to a model. In feature selection, it is this group of variables that we wish to reduce in size. Output variables are those for which a model is intended to predict, often called the response variable.

[Start Machine Learning](#)

The type of response variable typically indicates the type of predictive modeling problem being performed. For example, a numerical output variable indicates a regression predictive modeling problem, and a categorical output variable indicates a classification predictive modeling problem.

- **Numerical Output:** Regression predictive modeling problem.
- **Categorical Output:** Classification predictive modeling problem.

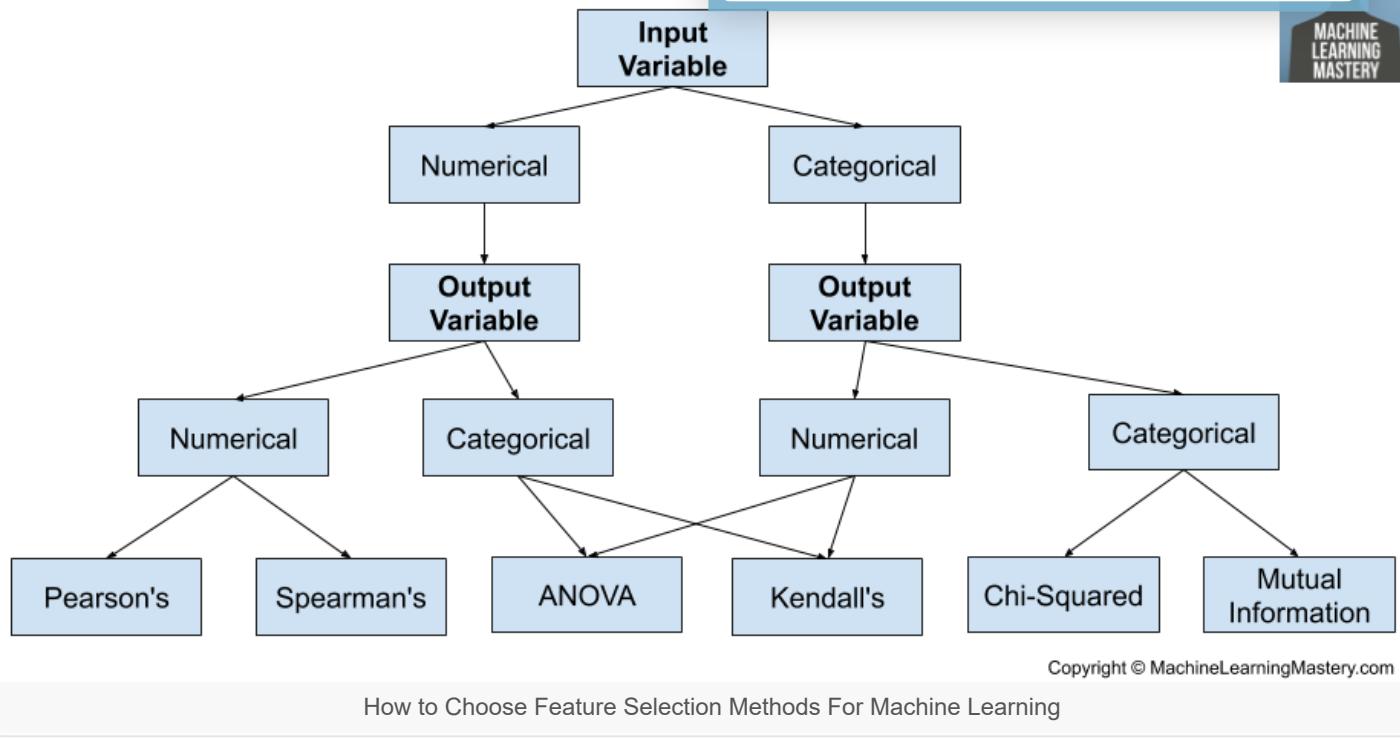
The statistical measures used in filter-based feature selection are generally calculated one input variable at a time with the target variable. As such, they are referred to as univariate statistical measures. This may mean that any interaction between input variables is not considered in the filtering process.

“Most of these techniques are univariate, meaning that they evaluate each predictor in isolation. In this case, the existence of correlated predictors can lead to redundant, predictors. The obvious consequence is that the same predictor can be chosen and, as a result, collinearity problems are often ignored.”

— Page 499, [Applied Predictive Modeling](#), 2013.

With this framework, let's review some univariate statistical methods for feature selection.

How to Choose a Feature Selection Method



Numerical Input, Numerical Output

This is a regression predictive modeling problem with numerical input variables.

[Start Machine Learning](#)

The most common techniques are to use a correlation coefficient, such as Pearson's for a linear correlation, or rank-based methods for a nonlinear correlation.

- Pearson's correlation coefficient (linear).
- Spearman's rank coefficient (nonlinear)

Numerical Input, Categorical Output

This is a classification predictive modeling problem with numerical input variables.

This might be the most common example of a classification problem,

Again, the most common techniques are correlation based, although in this case, they must take the categorical target into account.

- ANOVA correlation coefficient (linear).
- Kendall's rank coefficient (nonlinear).

Kendall does assume that the categorical variable is c

Categorical Input, Numerical Output

This is a regression predictive modeling problem with c

This is a strange example of a regression problem (e.g.,

Nevertheless, you can use the same “*Numerical Input, Categorical Output*” methods (described above), but in reverse.

Categorical Input, Categorical Output

This is a classification predictive modeling problem with categorical input variables.

The most common correlation measure for categorical data is the [chi-squared test](#). You can also use mutual information (information gain) from the field of information theory.

- Chi-Squared test (contingency tables).
- Mutual Information.

In fact, mutual information is a powerful method that may prove useful for both categorical and numerical data, e.g. it is agnostic to the data types.

3. Tips and Tricks for Feature Selection

This section provides some additional considerations when using filter-based feature selection.

Correlation Statistics

[Start Machine Learning](#)

The scikit-learn library provides an implementation of most of the useful statistical measures.

For example:

- Pearson's Correlation Coefficient: [f_regression\(\)](#)
- ANOVA: [f_classif\(\)](#)
- Chi-Squared: [chi2\(\)](#)
- Mutual Information: [mutual_info_classif\(\)](#) and [mutual_info_regression\(\)](#)

Also, the SciPy library provides an implementation of many more statistics, such as Kendall's tau ([kendalltau](#)) and Spearman's rank correlation ([spearmanr](#)).

Selection Method

The scikit-learn library also provides many different filter methods for each input variable with the target.

Two of the more popular methods include:

- Select the top k variables: [SelectKBest](#)
- Select the top percentile variables: [SelectPercentile](#)

I often use [SelectKBest](#) myself.

Start Machine Learning

You can master applied Machine Learning without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

Transform Variables

Consider transforming the variables in order to access different statistical methods.

For example, you can transform a categorical variable to ordinal, even if it is not, and see if any interesting results come out.

You can also make a numerical variable discrete (e.g. bins); try categorical-based measures.

Some statistical measures assume properties of the variables, such as Pearson's that assumes a Gaussian probability distribution to the observations and a linear relationship. You can transform the data to meet the expectations of the test and try the test regardless of the expectations and compare results.

What Is the Best Method?

There is no best feature selection method.

Just like there is no best set of input variables or best machine learning algorithm. At least not universally.

Instead, you must discover what works best for your specific problem using careful systematic experimentation.

Try a range of different models fit on different subsets of features chosen via different statistical measures and discover what works best for your specific problem.

4. Worked Examples of Feature Selection

It can be helpful to have some worked examples that you can copy-and-paste and adapt for your own project.

This section provides worked examples of feature selection cases that you can use as a starting point.

Regression Feature Selection: (Numerical Input, Numerical Output)

This section demonstrates feature selection for a regression problem that has numerical inputs and numerical outputs.

A test regression problem is prepared using the [make_regression\(\)](#) function.

Feature selection is performed using [Pearson's Correlation](#).

```

1 # pearson's correlation feature selection for
2 from sklearn.datasets import make_regression
3 from sklearn.feature_selection import SelectKBest
4 from sklearn.feature_selection import f_regression
5 # generate dataset
6 X, y = make_regression(n_samples=100, n_features=10)
7 # define feature selection
8 fs = SelectKBest(score_func=f_regression, k=10)
9 # apply feature selection
10 X_selected = fs.fit_transform(X, y)
11 print(X_selected.shape)

```

Start Machine Learning

You can master applied Machine Learning without math or fancy degrees.

Find out how in this *free* and *practical* course.

START MY EMAIL COURSE

Running the example first creates the regression dataset, then defines the feature selection and applies the feature selection procedure to the dataset, returning a subset of the selected input features.

```
1 (100, 10)
```

Classification Feature Selection: (Numerical Input, Categorical Output)

This section demonstrates feature selection for a classification problem that has numerical inputs and categorical outputs.

A test regression problem is prepared using the [make_classification\(\)](#) function.

Feature selection is performed using [ANOVA F measure](#) via the [f_classif\(\)](#) function.

```

1 # ANOVA feature selection for numeric input and categorical output
2 from sklearn.datasets import make_classification
3 from sklearn.feature_selection import SelectKBest
4 from sklearn.feature_selection import f_classif
5 # generate dataset
6 X, y = make_classification(n_samples=100, n_features=20, n_informative=2)
7 # define feature selection
8 fs = SelectKBest(score_func=f_classif, k=2)

```

Start Machine Learning

```

9 # apply feature selection
10 X_selected = fs.fit_transform(X, y)
11 print(X_selected.shape)

```

Running the example first creates the classification dataset, then defines the feature selection and applies the feature selection procedure to the dataset, returning a subset of the selected input features.

```
1 (100, 2)
```

Classification Feature Selection: (Categorical Input, Categorical Output)

For examples of feature selection with categorical inputs and categorical outputs, see the tutorial:

- [How to Perform Feature Selection with Categorical Inputs and Outputs](#)

Further Reading

This section provides more resources on the topic if you are interested.

Tutorials

- [How to Calculate Nonparametric Rank Correlation](#)
- [How to Calculate Correlation Between Variables in Python](#)
- [Feature Selection For Machine Learning in Python](#)
- [An Introduction to Feature Selection](#)

Start Machine Learning

You can master applied Machine Learning without math or fancy degrees.

Find out how in this *free* and *practical* course.

[START MY EMAIL COURSE](#)

Books

- [Applied Predictive Modeling](#), 2013.
- [Feature Engineering and Selection](#), 2019.
- [Feature Engineering for Machine Learning](#), 2018.

Articles

- [Feature selection, scikit-learn API.](#)
- [What are the feature selection options for categorical data? Quora.](#)

Summary

In this post, you discovered how to choose statistical measures for filter-based feature selection with numerical and categorical data.

Specifically, you learned:

- There are two main types of feature selection techniques: supervised and unsupervised, and supervised methods may be divided into wrapper, filter and intrinsic.

[Start Machine Learning](#)

- Filter-based feature selection methods use statistical measures to score the correlation or dependence between input variables that can be filtered to choose the most relevant features.
- Statistical measures for feature selection must be carefully chosen based on the data type of the input variable and the output or response variable.

Do you have any questions?

Ask your questions in the comments below and I will do my best to answer.

 Tweet

 Share

 Share



About Jason Brownlee

Jason Brownlee, PhD is a machine learning specialist who teaches developers how to get results with modern machine learning methods via hands-on practice.

[View all posts by Jason Brownlee →](#)

< How to Perform Feature Selection with Categorical Data

Start Machine Learning

You can master applied Machine Learning without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE

81 Responses to *How to Choose a Feature Selection Method For Machine Learning*



Mehmet F Yildirim November 27, 2019 at 7:19 am #

REPLY ↗

Hi Jason,

Thank you for the nice blog. Do you have a summary of unsupervised feature selection methods?



Jason Brownlee November 27, 2019 at 7:35 am #

REPLY ↗

All of the statistical methods listed in the post are unsupervised.



Markus November 28, 2019 at 12:43 am #

REPLY ↗

Hi

Actually I've got the same question like Mehmet above. Please correct me if I'm wrong, the talk in this article is about input variables and target variables. With that I understand features and labels of a given supervised learning problem.

Start Machine Learning

But in your answer it says unsupervised! I'm a bit confused.

Thanks.



Jason Brownlee November 28, 2019 at 6:41 am #

REPLY ↗

They are statistical tests applied to two variables, there is no supervised learning model involved.

I think by unsupervised you mean no target variable. In that case you cannot do feature selection. But you can do other things, like dimensionality reduction, e.g. SVM and PCA.



Markus November 28, 2019 at 7:00 am #

But the two code samples you provided are for supervised learning:

- Regression Feature Selection (Numerical)
- Classification Feature Selection (Numerical)

Do you maybe mean that supervised learning is the type of learning used for feature selection BUT this is not necessarily the case.

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee November 29, 2019 at 6:46 am #

Perhaps I am saying that this type of feature selection only makes sense on supervised learning, but it is not a supervised learning type algorithm – the procedure is applied in an unsupervised manner.



Markus November 30, 2019 at 3:19 am #

OK I guess know I understand what you mean.

Feature selection methods are used by the supervised learning problems to reduce the number of input features (or as you call them “the input variables”), however ALL of these methods themselves work in an unsupervised manner to do so.



Jason Brownlee November 30, 2019 at 6:30 am #

That is my understanding.

Start Machine Learning



What do you mean by unsupervised – like feature selection for clustering?



Jayant Vyas May 14, 2020 at 4:48 am #

Hello Sir,

If we have no target variable, can we apply feature selection before the clustering of a numerical dataset?



Jason Brownlee May 14, 2020

No. Feature selection requires

You can use unsupervised methods to this yet.



RajA November 27, 2019 at 3:39 pm #

Thanks again for short and excellent post. How also be used to identify best features.

Start Machine Learning

You can master applied Machine Learning without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee November 28, 2019 at 6:32 am #

REPLY ↗

Yes, but in this post we are focused on univariate statistical methods, so-called filter feature selection methods.



RajA November 28, 2019 at 5:51 pm #

REPLY ↗

Thanks for your time for the clarification.



Saurabh December 4, 2019 at 2:58 am #

REPLY ↗

Thanks for sharing. Actually I was looking for such a great blog since a long time.

Jason Brownlee December 4, 2019 at 5:45 am # [Start Machine Learning](#)

REPLY ↗



Thanks!

I hope it helps.



Abnádia Lura December 10, 2019 at 12:31 am #

REPLY ↵

Please give two reasons why it may be desirable to perform feature selection in connection with document classification.



Jason Brownlee December 10, 2019 at 7:34 am #

What would feature selection for documents do in reducing the size of the vocab?



Ayushi Verma December 10, 2019 at 8:57 pm #

quite an informative article with great content

Start Machine Learning

You can master applied Machine Learning without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee December 11, 2019 at 6:52 am #

REPLY ↵

Thanks!



YXZ December 18, 2019 at 2:00 am #

REPLY ↵

Hi Jason! Thanks for this informative post! I'm trying to apply this knowledge to the Housing Price prediction problem where the regressors include both numeric features and categorical features. In your graph, (Categorical Inputs, Numerical Output) also points to ANOVA. To use ANOVA correctly in this Housing Price case, do I have to encode my Categorical Inputs before SelectKBest?



Jason Brownlee December 18, 2019 at 6:09 am #

REPLY ↵

Yes, categorical variables will need to be label/integer encoded at the least.



Himanshu January 7, 2020 at 12:25 am #

Start Machine Learning

REPLY ↵

^

Hi Jason! I have dataset with both numerical and categorical features. The label is categorical in nature. Which is the best possible approach to find feature importance? Should I OneHotEncode my categorical features before applying ANOVA/Kendall's?



Jason Brownlee January 7, 2020 at 7:23 am #

REPLY ↗

Use separate statistical feature selection methods for different variable types.

Or try RFE.



Abhay January 12, 2020 at 12:44 am #

Hey Jason,

Thanks a lot for this detailed article.

I have a question, after one hot encoding my categorical output variable is numerical and all other predictors are numerical. Correlation for feature selection here (and for removing

Now since one hot encoded column has some ordinality, correlation matrix will be useful.

I tried this and the output is making sense business wise. Just wanted to know your thoughts on this, is this fundamentally correct ??

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee January 12, 2020 at 8:04 am #

REPLY ↗

No, spearman/pearson correlation on binary attributes does not make sense.

You perform feature selection on the categorical variables directly.



Zahra January 22, 2020 at 10:51 pm #

REPLY ↗

Thanks a lot for your nice post. I'm way new to ML so I have a really rudimentary question. Suppose I have a set of tweets which labeled as negative and positive. I want to perform some sentiment analysis. I extracted 3 basic features: 1. Emotion icons 2.Exclamation marks 3. Intensity words(very, really). My question is: How should I use these features with SVM or other ML algorithms? In other words, how should I apply the extracted features in SVM algorithm? Should I train my dataset each time with one feature? I read several articles and they are just saying: we should extract features and deploy them in our algorithms but HOW? Help me, please

Start Machine Learning



Jason Brownlee January 23, 2020 at 6:34 am #

REPLY ↤

The text will need a numeric representation, such as a bag of words.

This is called natural language processing, you can get started here:

<https://machinelearningmastery.com/start-here/#nlp>



Zahra January 23, 2020 at 8:10 am #

REPLY ↤

Thank you sir



Jason Brownlee January 23, 2020 #

You're welcome.



Sam January 25, 2020 at 4:15 am #

Hey jason,

Can you please say why should we use univariate selection method for feature selection?

Cause we should use correlation matrix which gives correlation between each dependent feature and independent feature,as well as correlation between two independent features.

So, using correlation matrix we can remove collinear or redundant features also.

So can you please say when should we use univariate selection over correlation matrix?

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee January 25, 2020 at 8:44 am #

REPLY ↤

Yes, filter methods like statistical test are fast and easy to test.

You can move on to wrapper methods like RFE later.



shadia January 25, 2020 at 5:08 am #

REPLY ↤

hi Jason

thnx for your helpful post

i want to know which method to use?

input vairables are

1. age

2.sex(but it has numbers as 1 for males and 2 for females)

Start Machine Learning

^

3. working hours-
 4. school attainment (also numbers)
 the output is numeric
 could you plz help



Jason Brownlee January 25, 2020 at 8:45 am #

REPLY ↗

Perhaps note whether each variable is numeric or categorical then follow the above guide.



Verdict February 25, 2020 at 5:33 am #

Hi, Jason!

Do you mean you need to perform feature selection based on parameters as illustrated above? Is there any specific way to calculate feature scores without worrying on the type of variable?



Jason Brownlee February 25, 2020

Yes, different feature selection for different variable types.

A short cut would be to use a different approach, like RFE, or an algorithm that does feature selection for you like xgboost/random forest.

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Charlotte January 31, 2020 at 3:08 pm #

REPLY ↗

Hello Jason,

Thank you for your nice blogs, I read several and find them truly helpful.

I have a quick question related to feature selection:
 if I want to select some features via VarianceThreshold, does this method only apply to numerical inputs?
 Can I encode categorical inputs and apply VarianceThreshold to them as well?

Many thanks!



Jason Brownlee February 1, 2020 at 5:46 am #

REPLY ↗

Thanks!

Yes, numerical only as far as I would expect.

Start Machine Learning



Tanuj February 21, 2020 at 3:32 am #

REPLY ↗

Hi Jason!

Is there any way to display the names of the features that were selected by SelectKBest?
In your example it just returns a numpy array with no column names.



Jason Brownlee February 21, 2020 at 8:29 am #

REPLY ↗

Yes, you can loop through the list of columns to see if they were selected or not using information from the attribute.

Start Machine Learning

You can master applied Machine Learning without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Sam February 22, 2020 at 3:24 am #

REPLY ↗

Hi Jason,

Many thanks for this detailed blog. A quick question on

Why do we select feature with high F value? Say if y = 1 because it has high F-statistic in a univariate ANOVA vs y = 0, is statistically different from the mean of feature 1 when y = 1? and therefore feature 1 likely to be useful in predicting y?



Jason Brownlee February 22, 2020 at 6:32 am #

REPLY ↗

Yes, large values. But don't do it manually use a built-in selection method.

See the worked examples at the end of the tutorial as a template.



bahri March 21, 2020 at 11:18 pm #

REPLY ↗

hi jason,

so im working with more than 100 thousand samples dota2 dataset which consist of the winner and the "hero" composition from each match. I was trying to build winner of the match prediction model similiar to this [<http://jmcauley.ucsd.edu/cse255/projects/fa15/018.pdf>]. so the vector input is

$X_i = 1$ if hero i on radiant side, 0 otherwise.

$X(119+i) = 1$ if hero i on dire side, 0 otherwise

Start Machine Learning

^

The vector X consist 238 entri since there are 119 kind of heroes. Each vector represent the composition of the heroes that is played within each match. Each match always consist of exactly 10 heroes (5 radiant side 5 dire side).

From this set up i would have a binary matrix of 100k times (222 + 1) dimension with row represent how many samples and columns represent features, +1 columns for the label vector (0 and 1, 1 meaning radiant side win)

so if i dot product between two column vector of my matrix, i can get how many times hero i played with hero j on all the samples.

so if i hadamard product between two column vector of my matrix and the result of that we dot product to the vector column label i can get how many times hero i played with hero j and win.

with this i can calculate the total weight of each entri per hero. I could get very high coorelation between this "new feature" to this problem in statistics textbook on binary data.



Jason Brownlee March 22, 2020 at 6:55 am #

Not sure I can offer good advice off the cuff.



khagesh March 24, 2020 at 4:36 am #

Hi Jason, Thanks for this article. I totally understand this different methodologies. I have one question.

If lets say. I have 3 variables. X,Y,Z

X= categorical

Y= Numerical

Z= Categorical, Dependent(Value I want to predict)

Now, I did not get any relationship between Y and Z and I got the Relationship between Y and Z. Is it possible that if we include X, Y both together to predict Z, Y might get the relationship with Z.

If is there any statistical method or research around please do mention them. Thanks



Jason Brownlee March 24, 2020 at 6:10 am #

REPLY ↗

I would recommend simply testing reach combination of input variables and use the combination that results in the best performance for predicting the target – it's a lot simpler than multivariate statistics.

[Start Machine Learning](#)

**San** March 31, 2020 at 8:29 am #

REPLY ↗

When having a dataset that contains only categorical variables including nominal, ordinal & dichotomous variables, is it incorrect if I use either Cramér's V or Theil's U (Uncertainty Coefficient) to get the correlation between features?

Thanks

San

**Jason Brownlee** March 31, 2020 at 8:38 am #

REPLY ↗

I don't know off-hand, sorry.

**Iraj** March 31, 2020 at 10:16 am #

REPLY ↗

Very good article.

I have detected outliers and wondering how can I estimate

We are talking about only one observation and it's labeled.
I couldn't find any reference for that.

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

START MY EMAIL COURSE**Jason Brownlee** March 31, 2020 at 1:36 pm #

REPLY ↗

This sounds like an open question.

Perhaps explore distance measures from a centroid or to inliers?

Or univariate distribution measures for each feature?

**Iraj** March 31, 2020 at 2:35 pm #

REPLY ↗

Thank you for quick response.

That's one class multivariate application.

For a single observation, I need to find out the first n features that have the most impact on being in that class.

From most articles, I can find the most important features over all observations, but here I need to know that over a selected observations.

Jason Brownlee April 1, 2020 at 5:45 am #**Start Machine Learning**

REPLY ↗



Simply fit the model on your subset of instances.



Suraj April 1, 2020 at 8:31 am #

REPLY ↗

- 1) In case of feature selection algorithm (XGBosst, GA, and PCA) what kind of method we can consider wrapper or filter?
- 2) what is the difference between feature selection and dimension reduction?



Jason Brownlee April 1, 2020 at 1:

XGBoost would be used as a filter selection method.

Feature selection chooses features in the original space. Dimension reduction projects the features into lower dimensional space.

Technically deleting features could be considered as a filter selection method.

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Suraj April 1, 2020 at 10:23 pm #

Thank you so much for your time to respond. Would you like to share some of the material on the same (so I can use it for my thesis as a reference)?

In addition, I am excited to know the advantages and disadvantaged in this respect; I mean when I use XGBoost as a filter feature selection and GA as a wrapper feature selection and PCA as a dimensional reduction, Then what may be the possible advantages and disadvantages?

best regards!



Jason Brownlee April 2, 2020 at 5:53 am #

If you need theory of feature selection, I recommend performing a literature review.

I cannot help you with advantages/disadvantages – it's mostly a waste of time. I recommend using what "does" work best on a specific dataset, not what "might" work best.



Iraj April 1, 2020 at 12:57 pm #

REPLY ↗

Start Machine Learning

I didn't get your point.

I have 1 record which is outlier. and wanted to know which features had the most contribution on that record to get outlier.

Thank you and sorry if question is confusing



Jason Brownlee April 1, 2020 at 1:35 pm #

REPLY ↗

I suggested that it is an open question – as in, there are no obvious answers.

I suggested to take it on as a research project and discover what works best.

Does that help, which part is confusing – perhaps I can elaborate?



Iraj April 1, 2020 at 2:13 pm #

Thank you

Should research on that



Jason Brownlee April 2, 2020 at 5:41 am #

Thanks for the suggestion.

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



CC April 9, 2020 at 6:14 pm #

REPLY ↗

Do the above algorithms keep track of 'which' features have been selected, or only selects the 'best' feature data? After having identified the 'best k features', how do we extract those features, ideally only those, from new inputs?



Jason Brownlee April 10, 2020 at 8:24 am #

REPLY ↗

Yes, you can discover which features are selected according to their column index.



Oliver Tomic April 14, 2020 at 12:14 am #

REPLY ↗

Hi Jason,

thanks again you for the nice overview. Here is relatively new package introducing the Phi_K correlation coefficient that claims that it can be used across categorical, ordinal and numerical features. It is also said to

capture non-linear dependency.

<https://phik.readthedocs.io/en/latest/#>

best wishes

Oliver



Jason Brownlee April 14, 2020 at 6:20 am #

REPLY ↗

Thanks for sharing.



Ahmed Jyad April 14, 2020 at 2:15 am #

Hi Jason, Thanks for the amazing article.
A question on using ANOVA. Given Categorical variables, we assume homogeneity of variance between the samples.
ANOVA require the assumption of equal variance.



Jason Brownlee April 14, 2020 at 6:26 am #

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Perhaps.

Often the methods fail gracefully rather than abruptly, which means you can use them reliably when assumptions are violated.



Recep April 18, 2020 at 5:21 am #

REPLY ↗

Hello Jason,

Thank you for your explanation and for sharing great articles with us!

You have clearly explained how to perform feature selection in different variations in the “How to Choose Feature Selection Methods For Machine Learning” table.

+ Numerical Input, Numerical Output:

Pearson’s correlation coefficient (linear). Spearman’s rank coefficient (nonlinear)

+ Numerical Input, Categorical Output:

ANOVA correlation coefficient (linear). Kendall’s rank coefficient (nonlinear).

+ Categorical Input, Numerical Output:

ANOVA correlation coefficient (linear). Kendall’s rank coefficient (nonlinear).

Start Machine Learning

^

+ Categorical Input, Categorical Output:

Chi-Squared test (contingency tables). Mutual Information.

I would like to ask some questions about the dataset that contains a combination of numerical and categorical inputs.

1- Which methods should we apply when we have a dataset that has a combination of numerical and categorical inputs? (e.g: Total Input: 50; Numerical:25 and Categorical:25. Task: Classification problem with the categorical values)

2- Should I apply one of the label encoding methods (encoding depending on the labels in the feature let's say I applied one-hot, target encoding). Get the numerical values from the categorical input. Then, my problem becomes into the Numerical Input, Categorical Output. In this time, should I apply to the ANOVA correlation coefficient (linear) and Kendall's rank coefficient (nonlinear) techniques?

3- OR, What would be the better approaches to apply to a Numerical Input, Categorical Output (Categorical Output) problem that includes a combination of numerical and categorical inputs?

Thank you.



Jason Brownlee April 18, 2020 at 6:13 am #

You're welcome.

You would use a separate method for each data type. For example, if you had a categorical variable with three categories, you would use a different method than if you had a numerical variable with three values.

Start Machine Learning

You can master applied Machine Learning without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Peter April 27, 2020 at 3:52 pm #

REPLY ↗

I have a mixture of numeric, ordinal, and nominal attributes.

For the first two, Pearson is used to determine the correlation with the target.

For the nominal type, I still cannot find a good reference on how we should handle it for correlation. Encode it to numeric doesn't seem correct as the numeric values would probably suggest some ordinal relationship but it should not for nominal attributes.

Any advice?

I also tested the model performance based on the transformed attribute that gives higher correlation with the target, but however, the model performance did not improve as expected. any suggestions what can i explore further?



Jason Brownlee April 28, 2020 at 6:43 am #

REPLY ↗

Nominal is "Categorical" now follow the above advice based on the type of the output variable.

Start Machine Learning



Andrew April 30, 2020 at 12:21 am #

REPLY ↗

Hi Jason, when the output, i.e. the label is 0 or 1 meant to represent Bad and Good, is this considered as numeric output or categorical output?



Jason Brownlee April 30, 2020 at 6:48 am #

REPLY ↗

It is a number that we can map to a category in our application.



VASHIST NARAYAN Singh May 9, 2020 at 1:48 pm #

REPLY ↗

Jason you have not shown the example of cat



Jason Brownlee May 9, 2020 at 1:52 pm #

REPLY ↗

You can reverse the case for: Numerical In

Start Machine Learning

You can master applied Machine Learning without math or fancy degrees.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Thaung Myint Htun May 9, 2020 at 5:50 pm #

REPLY ↗

Sir,

In my dataset, 29 attributes are yes/no values(binary) and the rest is numeric(float)type attributes. Class has 7 values(multiclass). I want to try this dataset for classification. Which techniques of feature selections are suitable? Please give me a hand!

With respect!

Thaung



Jason Brownlee May 10, 2020 at 6:00 am #

REPLY ↗

Some ideas:

Perhaps establish a baseline performance with all features?

Perhaps try separate feature selection methods for each input type?

Perhaps try a wrapper method like RFE that is agnostic to input type?



Zineb May 11, 2020 at 11:06 am #

Start Machine Learning

REPLY ↗



Hi Jason and thanks a lot for this wonderful and so helpful work.

1- wrapper methods : does the model get rid of irrelevant features or it just assigns small weights.?

2-filter methods: when using univariate techniques, how to detect or prevent redundant among the selected variables ?

3- What about t-statistic ?

Thanks in advance.



Jason Brownlee May 11, 2020 at 1:36 pm #

REPLY ↗

Wrapper tests different subsets of features and chooses a subset that gives best skill.

Deleting redundant features is performed without them dropped.

Yes, I have read this. I don't have an example.



Zineb May 12, 2020 at 6:00 am #

REPLY ↗

Thanks a lot.

Start Machine Learning

You can master applied Machine Learning **without math or fancy degrees**.

Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Jason Brownlee May 12, 2020 at 6:54 am #

REPLY ↗

You're welcome.



Jayant Vyas May 13, 2020 at 11:08 pm #

REPLY ↗

Hello Sir,

I am new to this subject, I want to apply UNSUPERVISED MTL NN model for prediction on a dataset, for that I have to first apply clustering to get the target value. I want to apply some feature selection methods for the better result of clustering as well as MTL NN methods, which are the feature selection methods I can apply on my numerical dataset.



Jason Brownlee May 14, 2020 at 5:51 am #

REPLY ↗

The above tutorial explains exactly, perhaps re-read it?

Start Machine Learning

^

Leave a Reply

Name (required)

Email (will not be published) (requi

Website

SUBMIT COMMENT

Start Machine Learning

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

Email Address

START MY EMAIL COURSE



Welcome!

My name is *Jason Brownlee* PhD, and I **help developers** get results with **machine learning**.
[Read more](#)

Never miss a tutorial:



Picked for you:



[Why One-Hot Encode Data in Machine Learning?](#)

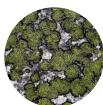


[An Introduction to Feature Selection](#)



[How to Prepare Data For Machine Learning](#)

Start Machine Learning

[How to Handle Missing Data with Python](#)[Discover Feature Engineering, How to Engineer Features and How to Get Good at It](#)

Loving the Tutorials?

The [EBook Catalog](#) is where I
keep the *Really Good* stuff.

[SEE WHAT'S NEW](#)

Start Machine Learning X

You can master applied Machine Learning
without math or fancy degrees.
Find out how in this *free* and *practical* course.

[START MY EMAIL COURSE](#)

© 2019 Machine Learning Mastery Pty. Ltd. All Rights Reserved.
Address: PO Box 206, Vermont Victoria 3133, Australia. | ACN 624 719 113
[LinkedIn](#) | [Twitter](#) | [Facebook](#) | [Newsletter](#) | [RSS](#)

[Privacy](#) | [Disclaimer](#) | [Terms](#) | [Contact](#) | [Sitemap](#) | [Search](#)

[Start Machine Learning](#)